

A New Generalization of Chebyshev Inequality for Random Vectors *

Xinjia Chen

June 2007

Abstract

In this article, we derive a new generalization of Chebyshev inequality for random vectors. We demonstrate that the new generalization is much less conservative than the classical generalization.

1 Classical Generalization of Chebyshev inequality

The Chebyshev inequality discloses the fundamental relationship between the mean and variance of a random variable. Extensive research works have been devoted to its generalizations for random vectors. For example, various generalizations can be found in Marshall and Olkin (1960), Godwin (1955), Mallows (1956) and the references therein. A natural generalization of Chebyshev inequality is as follows.

For a random vector $X \in \mathbf{R}^n$ with cumulative distribution $F(\cdot)$,

$$\Pr \{ \|X - E[X]\| \geq \varepsilon \} \leq \frac{\text{Var}(X)}{\varepsilon^2} \quad \forall \varepsilon > 0 \quad (1)$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector and

$$\text{Var}(X) \stackrel{\text{def}}{=} \int_{V \in \mathbf{R}^n} \|V - E[X]\|^2 dF(V)$$

This classical generalization can be found in a number of textbooks of probability theory and statistics (see, e.g., pp. 446-451 of Laha and Rohatgi (1979)).

*The author is with Department of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA 70803; Email: chenxinjia@gmail.com.

2 New Generalization of Chebyshev inequality

The classical generalization (1) perfectly assembles its counterpart for scalar random variables. However, it may be too conservative. To address the conservatism, we derive a new multivariate Chebyshev inequality as follows.

Theorem 1 For any random vector $X \in \mathbf{R}^n$ with covariance matrix Σ ,

$$\Pr \left\{ (X - E[X])^\top \Sigma^{-1} (X - E[X]) \geq \varepsilon \right\} \leq \frac{n}{\varepsilon}, \quad \forall \varepsilon > 0 \quad (2)$$

where the superscript “ \top ” denotes the transpose of a matrix.

Proof. Let $\mathbf{D}_\varepsilon = \{V \in \mathbf{R}^n : (V - E[X])^\top \Sigma^{-1} (V - E[X]) \geq \varepsilon\}$. By the definition of \mathbf{D}_ε , we have

$$\frac{1}{\varepsilon} (V - E[X])^\top \Sigma^{-1} (V - E[X]) \geq 1, \quad \forall V \in \mathbf{D}_\varepsilon.$$

Hence,

$$\begin{aligned} \Pr \{X \in \mathbf{D}_\varepsilon\} &\leq \frac{1}{\varepsilon} \int_{V \in \mathbf{D}_\varepsilon} (V - E[X])^\top \Sigma^{-1} (V - E[X]) dF(V) \\ &\leq \frac{1}{\varepsilon} \int_{V \in \mathbf{R}^n} (V - E[X])^\top \Sigma^{-1} (V - E[X]) dF(V). \end{aligned}$$

For $i = 1, \dots, n$, let u_i denote the i -th element of $V - E[X]$. For $i = 1, \dots, n$ and $j = 1, \dots, n$, let σ_{ij} denote the element of Σ in the i -th row and j -th column. Similarly, let ρ_{ij} denote the element of Σ^{-1} in the i -th row and j -th column. Then,

$$\begin{aligned} (V - E[X])^\top \Sigma^{-1} (V - E[X]) &= \sum_{i=1}^n u_i \left(\sum_{k=1}^n \rho_{ik} u_k \right) \\ &= \sum_{i=1}^n \sum_{k=1}^n \rho_{ik} u_i u_k. \end{aligned}$$

It follows that

$$\begin{aligned} &\int_{V \in \mathbf{R}^n} (V - E[X])^\top \Sigma^{-1} (V - E[X]) dF(V) \\ &= \int_{V \in \mathbf{R}^n} \left(\sum_{i=1}^n \sum_{k=1}^n \rho_{ik} u_i u_k \right) dF(V) \\ &= \sum_{i=1}^n \sum_{k=1}^n \rho_{ik} \left[\int_{V \in \mathbf{R}^n} u_i u_k dF(V) \right]. \end{aligned}$$

By the definition of the covariance matrix Σ and its symmetry, we have

$$\int_{V \in \mathbf{R}^n} u_i u_k dF(V) = \sigma_{ik} = \sigma_{ki}$$

for $i = 1, \dots, n$ and $k = 1, \dots, n$. Hence,

$$\begin{aligned}
& \int_{V \in \mathbf{R}^n} (V - E[X])^\top \Sigma^{-1} (V - E[X]) dF(V) \\
&= \sum_{i=1}^n \sum_{k=1}^n \rho_{ik} \sigma_{ki} \\
&= \text{tr}(\Sigma^{-1} \Sigma) \\
&= n
\end{aligned}$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix. Therefore,

$$\begin{aligned}
\Pr \{X \in \mathbf{D}_\varepsilon\} &\geq \frac{1}{\varepsilon} \int_{V \in \mathbf{R}^n} \left(\Sigma^{-1} (V - E[X]) (V - E[X])^\top \right) dF(V) \\
&= \frac{n}{\varepsilon}.
\end{aligned}$$

The proof is thus completed. □

Remark 1 *Theorem 1 indicates a fundamental relationship between the mean and covariance of a random vector and describes how a random vector deviates from its expectation. Specially, for $n = 1$, we have $\Sigma = \text{Var}(X)$ and by Theorem 1, for any $\epsilon > 0$,*

$$\begin{aligned}
& \Pr \left\{ (X - E[X])^\top \Sigma^{-1} (X - E[X]) > \epsilon \right\} \\
&= \Pr \left\{ \|X - E[X]\| > \sqrt{\epsilon \text{Var}(X)} \right\} \\
&\leq \frac{1}{\epsilon},
\end{aligned}$$

from which we deduce

$$\Pr \{ \|X - E[X]\| > \varepsilon \} \leq \frac{\text{Var}(X)}{\varepsilon^2}$$

by letting $\varepsilon = \sqrt{\epsilon \text{Var}(X)}$. This shows that Theorem 1 includes the well-known Chebyshev inequality as a special case.

3 Comparison with Classical Generalization

In this section, we shall show that the inequality in Theorem 1 can be much less conservative than the classical generalized Chebyshev inequality (1).

Let $\delta \in (0, 1)$. Based on inequality (1), sphere

$$\mathbf{B}_\delta \stackrel{\text{def}}{=} \left\{ V \in \mathbf{R}^n : \|V - E[X]\|^2 \leq \frac{\text{tr}(\Sigma)}{\delta} \right\}$$

is the smallest set that can be constructed to ensure $\Pr\{X \in \mathbf{B}_\delta\} > 1 - \delta$. On the other hand, by applying Theorem 1 we can construct an ellipsoid

$$\mathbf{E}_\delta \stackrel{\text{def}}{=} \left\{ V \in \mathbf{R}^n : (V - E[X])^\top \Sigma^{-1} (V - E[X]) \leq \frac{n}{\delta} \right\},$$

which guarantees $\Pr\{X \in \mathbf{E}_\delta\} > 1 - \delta$.

For a comparison of the conservativeness of generalized Chebyshev inequalities (1) and (2), it is natural to consider the ratio $\frac{\text{vol}(\mathbf{B}_\delta)}{\text{vol}(\mathbf{E}_\delta)}$ where $\text{vol}(\cdot)$ is a volume function such that $\text{vol}(S) = \int_{v \in S} dv$ for any $S \subset \mathbf{R}^n$. Interestingly, we have

Theorem 2 *For any random vector $X \in \mathbf{R}^n$,*

$$\frac{\text{vol}(\mathbf{B}_\delta)}{\text{vol}(\mathbf{E}_\delta)} = \frac{\left(\sqrt{\frac{\text{tr}(\Sigma)}{n}} \right)^n}{\sqrt{\det(\Sigma)}} > 1$$

where $\det(\Sigma)$ is the determinant of Σ .

Proof. By the definitions of variance and covariance, we have $\text{Var}(X) = \text{tr}(\Sigma)$. It follows that

$$\text{vol}(\mathbf{B}_\delta) = K \left(\sqrt{\frac{\text{tr}(\Sigma)}{\delta}} \right)^n$$

where $K > 0$ is a constant. Applying a linear transform $u = \Sigma^{-\frac{1}{2}}(v - E[X])$ to the integration $\text{vol}(\mathbf{E}_\delta) = \int_{v \in \mathbf{E}_\delta} dv$, we have

$$\text{vol}(\mathbf{E}_\delta) = \det(\Sigma^{\frac{1}{2}}) \int_{\|u\|^2 \leq \frac{n}{\delta}} du = \sqrt{\det(\Sigma)} K \left(\sqrt{\frac{n}{\delta}} \right)^n$$

and thus

$$\frac{\text{vol}(\mathbf{B}_\delta)}{\text{vol}(\mathbf{E}_\delta)} = \frac{\left(\sqrt{\frac{\text{tr}(\Sigma)}{n}} \right)^n}{\sqrt{\det(\Sigma)}}.$$

To show $\frac{\text{vol}(\mathbf{B}_\delta)}{\text{vol}(\mathbf{E}_\delta)} > 1$, it is equivalent to show

$$\frac{\text{tr}(\Sigma)}{n} \geq [\det(\Sigma)]^{\frac{1}{n}}.$$

Recall that the geometric average is no less than the arithmetic average,

$$\frac{\text{tr}(\Sigma)}{n} = \frac{\sum_{i=1}^n \sigma_{ii}}{n} \geq \left(\prod_{i=1}^n \sigma_{ii} \right)^{\frac{1}{n}}, \quad (3)$$

where σ_{ii} , $i = 1, \dots, n$ are the diagonal components of Σ . Note that the covariance matrix Σ is positive definite, hence by Hadamard's inequality,

$$\det(\Sigma) \leq \prod_{i=1}^n \sigma_{ii}. \quad (4)$$

It follows from (3) and (4) that $\frac{\text{tr}(\Sigma)}{n} \geq [\det(\Sigma)]^{\frac{1}{n}}$. The proof is thus completed. \square

As an illustrative example, consider a two-dimensional random vector

$$X = \begin{bmatrix} y \\ y + z \end{bmatrix}$$

where y and z are independent Guassian random variables with zero means and variances σ^2 , $k\sigma^2$ respectively. Straightforward computation gives

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & (k+1)\sigma^2 \end{bmatrix}$$

and

$$\frac{\text{vol}(\mathbf{B}_\delta)}{\text{vol}(\mathbf{E}_\delta)} = \frac{k+2}{2\sqrt{k}} \geq \sqrt{2}.$$

Obviously, as k increases from 2 to ∞ or decreases from 2 to 0, the ratio of volumes increases monotonically and tends to ∞ .

In the following Figure 1, ellipsoid \mathbf{E}_δ and sphere \mathbf{B}_δ are constructed for $\sigma = 1$, $k = 25$ and $\delta = 0.1$. Moreover, 1000 i.i.d. samples of X are generated to show the coverage of the ellipsoid and sphere. It can be seen that most samples are included in the ellipsoid. This indicates that Theorem 1 is much less conservative than the classical generalized Chebyshev inequality in describing how a random vector deviates from its expectation.

References

- [1] H. J. Godwin, "On Generalizations of Chebyshev Inequality," *Journal of the American Statistical Association*, pp. 923-945, Vol. 50, No. 271, 1955.
- [2] R. G. Laha and V. K. Rohatgi, *Probability Theory*, pp. 446-451, John Wiley and Sons, 1979.
- [3] C. L. Mallows, "Generalizations of Chebyshev Inequalities," *Journal of the Royal Statistical Society, Series B*, pp. 139-171, Vol. 18, No. 2, 1956.
- [4] A. W. Marshall and I. Olkin, "Multivariate Chebyshev Inequalities," *The Annals of Mathematical Statistics*, pp. 1001-1014, Vol. 31, No. 4, 1960.

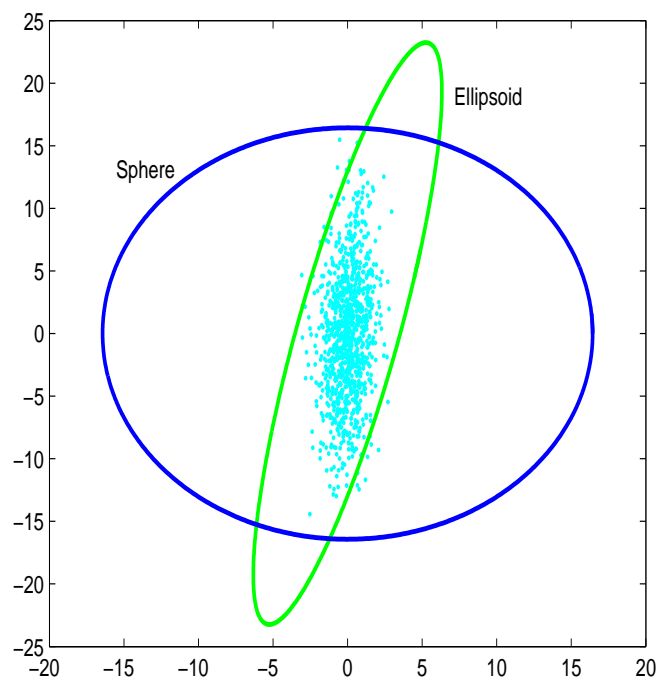


Figure 1: Comparison of Generalized Chebyshev Inequalities